DOCUMENT RESUME

ED 422 392 TM 028 954

AUTHOR	Zhang, Shuqiang
TITLE	Fourteen Homogeneity of Variance Tests: When and How To Use Them.
PUB DATE	1998-04-00
NOTE	18p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE	Information Analyses (070) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Software; *Statistical Analysis; *Test Use; Textbook Content
IDENTIFIERS	*Homogeneity of Variance

ABSTRACT

Homogeneity of variance (HOV) is a major assumption underlying the validity of many parametric tests. More importantly, it serves as the null hypothesis in substantive studies that focus on cross- or within-group dispersion. Despite a widely acknowledged need for testing HOV, very few textbooks give adequate coverage on the topic, and many HOV tests are still missing from statistical software packages. Using language comprehensible to those who have completed only 1 introductory statistics course in college, this paper explains 14 representative HOV tests for 5 types of research situations: (1) 1-sample HOV test; (2) 2-sample HOV test; (3) HOV test involving 2 or more samples; (4) HOV test for factorial designs; and (5) HOV tests for 2 correlated samples. Brief guidelines are provided as to when and how each of the HOV tests is to be used, and sample programs are included for HOV tests available from the SAS/STAT system. All the remaining tests can be very easily calculated by hand using descriptive statistics. The paper concludes with a conceptual summary of four major approaches to HOV testing. (Contains 19 references.) (SLD)

FOURTEEN HOMOGENEITY OF VARIANCE TESTS: WHEN AND HOW TO USE THEM

Shuqiang Zhang
University of Hawaii at Manoa

Presented
at the 1998 Annual Meeting of the
American Educational Research Association
April 13-17, 1998
San Diego, California

ABSTRACT

Homogeneity of variance (HOV) is a major assumption underlying the validity of many parametric tests. More importantly, it serves as the null hypothesis in substantive studies that focus on cross- or within-group dispersion. Despite a widely acknowledged need for testing HOV, very few textbooks give adequate coverage on the topic, and many HOV tests are still missing from statistical software packages.

Using language comprehensible to those who have completed only one introductory statistics course in college, this paper explains 14 representative HOV tests for five types of research situations. Brief guidelines are provided as to when and how each of the HOV tests is to be used, and sample programs are included for HOV tests available from SAS/STAT. All the remaining tests can be easily calculated by hand using descriptive statistics. The paper concludes with a conceptual summary of four major approaches to HOV testing.

Introduction

The statistical validity of many commonly used tests such as the t-test and ANOVA depends on the extent to which the data conform to the assumption of homogeneity of variance (HOV). When a research design involves groups that have very different variances, the p value accompanying the test statistics, such as t and F, may be too lenient or too harsh. Furthermore, substantive research often requires investigation of cross- or within-group fluctuation in dispersion. For example, in quality control research, HOV tests are often "a useful endpoint in an analysis" (Conover, Johnson & Johnson, 1981, p. 351). In human performance studies, an increase or decrease in the dispersion of performance scores within the same group of subjects may shed light on how changing conditions affect human behavior. Recent studies on gender-related differences in the dispersion of academic performance have provoked substantive as well as methodological interest in HOV (e.g., Feingold, 1992; Noddings, 1992; Shaffer, 1992; Hedges & Friedman, 1993). Gould (1996) recommends a close scrutiny of decreasing or increasing variation within a complex system for a more accurate interpretation of trends.

Despite an acknowledged need for testing HOV, such tests are seldom taught and often missing from software packages. This paper explains how 14 representative HOV tests may be performed for five types of research designs and concludes with a conceptual summary of four major approaches to HOV testing.

I. One-Sample HOV Test

A convenient **chi-square test** can determine whether the difference between a sample variance and a known or posited population variance is large enough to reject the null hypothesis, H_o: $\sigma_1^2 = \sigma_0^2$. SAS/STAT does not have a special procedure for the test. However, once S^2 is known through PROC MEANS or any option that provides basic descriptive statistics, the test can be done with minimal computation.

$$\chi^2 = (n-1)S^2 / \sigma_0^2$$

where n = sample size
S^2 = sample variance
σ_0^2 = population variance

The χ^2 test has (n-1) degrees of freedom. The critical value for a chosen significance level can be found in the χ^2 table available in most statistics textbooks. The test is not accurate when the population deviates from normality and the sample size is small.

II. Two-Sample HOV Test

This test, known as the **folded form F test**, is automatically conducted when PROC TTEST is invoked. The folded form F test uses the ratio of the larger variance to the smaller variance to test the null hypothesis, H_o: $\sigma_1^2 = \sigma_2^2$.

$$F' = S_l^2 / S_s^2$$

where S_l^2 = larger variance
S_s^2 = smaller variance

The following SAS statements, with GROUP as the independent variable and SCORE as the dependent variable, produce, among other things, the folded form F′:

```
PROC TTEST;
CLASS GROUP;
VAR SCORE;
RUN;
```

The test has $(n_l - 1)$ and $(n_s - 1)$ degrees of freedom for the numerator and the denominator respectively. Because the larger variance is always taken to be the numerator, F′ is always larger than 1. In other words, only one direction of the F distribution is considered. SAS/STAT adjusts for the directional tail and prints out the correct p value. Should anyone try to conduct the test by hand and refer to the conventional F table, he or she needs to remember that the listed critical F at the significance level of 0.05 actually means a significant level of approximately 0.10 in the case of the folded form F test (Ferguson, 1981, pp. 189-192). The test is very sensitive to deviations from the normal distribution.

III. HOV Tests Involving Two Or More Samples

Hartley's F_{max} test is a shortcut method for testing the overall null, $H_o: \sigma_1^2 = \sigma_2^2 \ldots = \sigma_k^2$. Instead of taking all the variances into account, it focuses only on the ratio between the largest and the smallest variance. It is the two-sample folded form F test generalized to more than two samples:

$$F_{\max} = S_{max}^{\;2} / S_{min}^{\;2}$$

where $S_{max}^{\;2}$ = maximum variance
$S_{min}^{\;2}$ = minimum variance

The test is not available from SAS and requires equal or roughly equal sample sizes under the assumption of normality. The table of critical F_{max} values for various combinations of k (number of groups) and n (if all the groups have the same size) can be found in Kanji (1993, p. 182) or Rosenthal and Rosnow (1992, pp. 608-609). When the groups have slightly different sample sizes, the harmonic mean may serve as the adjusted sample size n′ (Rosenthal & Rosnow, 1991, pp. 338-339):

$$\text{harmonic mean} = k / \sum n_j^{\;-1}$$

where k = number of groups
n_j = size of the j^th^ group

Even though the F_{max} test can reject the overall null, it cannot pinpoint between which two groups heterogeneity of variance occurs. For that purpose, the researcher needs a multiple test analogous to Duncan's test following the rejection of the overall null in one-way ANOVA. **David's multiple test** (1954) extends the folded form F test a step further to pairwise comparisons among k groups, always placing the larger variance over the smaller one, as is done in the folded form F′ formula. For critical values for the Duncan-type multiple HOV test, see Tietjen & Beckman's maximum F-ratio table (1972). This test requires equal or roughly equal group sizes and is very sensitive to departures from the normal distribution.

As an improvement on Hartley's F_{max} test, which involves only the maximum and the minimum variances, **Cochran's G test**, also known as Cochran's C test, uses the dispersion information in all the k groups. It is appropriate for equal or roughly equal size groups and is typically used in the situation where one group seems to be drastically more spread out than all the other groups sharing more or less the same variance. In that sense, it is a test to identify an outlier in terms of variance.

$$G = S_{max}^2 / (k\, MS_{error})$$

where S_{max}^2 = maximum variance
N = sum of all sample sizes
k = number of groups
$MS_{error} = \sum(X_{ij} - \overline{X}_j)^2 / (N-k)$

Cochran's G, or C, is basically a variance ratio, except that the denominator is the product of k (number of groups) and the pooled within-group variance, often referred to as MS_{within} or MS_{error}, available from the one-way ANOVA printout. SAS does not have an option for the test, but it can be done indirectly through ANOVA plus a little bit of calculation. The following SAS statement generates, among other things, MS_{error}:

```
PROC GLM;
CLASS GROUP;
MODEL SCORE = GROUP;
RUN;
```

This test requires a special table of critical values for various combinations of k and n (Rosenthal & Rosnow, 1991, pp. 610-611; Winer 1971, p. 876). The harmonic mean may be adopted as the adjusted n′ if the groups have roughly equal sizes.

Unlike all the previous tests that directly compare two variances in a ratio, the **Bartlett-Kendall test** uses the log transformation of the variance, because the sampling distribution of the log variance is normally distributed. The numerator in the formula is the log of a variance ratio.

$$Z_{B\text{-}K} = [\ln (S_{max}^2) - \ln (S_{min}^2)] / (n/2)^{0.5}$$

where n = sample size
S_{max}^2 = maximum variance
S_{min}^2 = minimum variance

This test applies to equal size samples. In case of samples with roughly equal sizes, the arithmetic average of the sample sizes is used in the formula. A special table is needed for critical values (Bartlett & Kendall, 1946; Pearson & Hartley, 1970, p. 203). The Bartlett-Kendall test and Hartley's F_{max} test, one using log transformation and the other using the variance ratio, produce practically identical results.

Another test that involves log transformation of the variance is the **Bartlett X^2 test**. The transformation allows the X^2 distribution to serve as the basis for rejection of the null. The log transformation also improves (though not much) robustness in case of departures from the normal distribution, but in doing so, reduces power slightly.

$$\chi^2 = \frac{N \cdot \ln \Sigma[\frac{(n_j - 1)}{N} S_j^2] - \ln \Sigma(n_j - 1)S_j^2}{1 + [(\Sigma \frac{1}{n_j - 1} - \frac{1}{N}) / 3(k-1)]}$$

where N = sum of all sample sizes
k = number of samples
n_j = size of the jth sample
S_j^2 = variance of the jth sample

The numerator is essentially based on the negative log of the ratio between the group variance and the geometric mean of the k group variances. The denominator is a correction factor to improve approximation to the χ^2 distribution. The chi-square test has (k-1) degrees of freedom. This likelihood test is sensitive to departures from the normal distribution. Preferably, the samples have comparable sizes. The Bartlett test does not have a subsequence multiple comparison procedure. The following SAS statements conduct the Bartlett test:

```
PROC GLM;
CLASS GROUP;
MODEL SCORE = GROUP;
MEANS GROUP / HOVTEST = BARTLETT;
RUN;
```

When the groups have different sizes, **Levene's test** is recommended. The test has two options. For Option One, group means are calculated first. For each person, the absolute deviation of the person's score from the mean of the group to which the person belongs is calculated, $|x_{ij} - x_j|$. This absolute deviation represents how far the person is displaced or spread out from the group mean. Such variables are known as spread or dispersion variables. Since the variance of each group is related to the sum of the absolute deviations within the group, testing the differences among the group means of the absolute deviations through the regular one-way ANOVA is tantamount to testing homogeneity of variance. Option One is recommended for highly skewed data. The SAS statements for Option One are included:

```
PROC GLM;
CLASS GROUP;
MODEL SCORE = GROUP;
MEANS GROUP / HOVTEST = LEVENE TYPE = ABS;
RUN;
```

Option Two shares the same logic with Option One, but the spread variable is the square of the absolute deviation. $(x_{ij} - x_j)^2$. SAS runs Option Two by default. One can also specify TYPE = SQUARE in the program above to call up Option Two. One weakness of Levene's test is that it may allow a higher Type I error rate than it should.

An improvement on Levene's test is the **Brown-Forsythe test**, which follows the same logic underlying one-way ANOVA except that the spread variable becomes the square of the deviation from the group median, rather than the group mean. When all the distributions are normal, the Brown-Forsythe test and the Levene's test Option Two are identical. The SAS Institute recommends the Brown-Forsythe test as the most powerful "to detect variance differences while protecting the Type I error probability" (1997, p. 356). It is not yet clear what multiple comparison options are appropriate following Levene's test or the Brown-Forsythe test.

Another ANOVA-based test is the **O'Brien test**, which relies on yet another spread variable r through a formula that allows the statistician to choose a weight (w) between 0 and 1 to adjust the transformation:

$$r_{ij} = [(w + n_j - 2)\, n_j (x_{ij} - \bar{x}_j)^2 - w\, S_j^2 (n_j - 1)] \,/\, [(n_j - 1)(n_j - 2)]$$

where w = weight (usually 0.5)
n_j = size of the j[th] group
S_j^2 = variance of the j[th] group
x_{ij} =score of the i[th] person in the j[th] group
$\bar{x}_j$ =mean of the j[th] group

The most commonly adopted w is 0.5 to offset the anticipated moderate departure from kurtosis=0. The actual kurtosis is almost never known, and the choice of w other than 0.5 rarely makes a critical difference in practice. When no w is specified, SAS, by default, coverts the dependent variable into r using w=0.5 and then subjects r to the regular one-way ANOVA. The following SAS statements accomplish the O'Brien test:

```
PROC GLM;
CLASS GROUP;
MODEL SCORE = GROUP;
MEANS GROUP / HOVTEST = OBRIEN W = 0.5;
RUN;
```

The W = 0.5 option above is redundant, but it demonstrates how the researcher can specify other values for the weight. O'Brien suggested a prudent procedure for subsequent contrasts (1981). Once the null H_0: $\sigma_1^2 = \sigma_2^2 \ldots = \sigma_k^2$ is rejected, the researcher need resort to Welch's variance-weighted one-way ANOVA (1951), which is robust to heterogeneity of variance:

$$F = \frac{[\Sigma w_j (\bar{x}_j - \bar{x}_{adj.})^2 / (k-1)]}{1 + \frac{2(k-1)}{(k^2-1)} [\Sigma \frac{(1 - \frac{w_j}{\Sigma w_j})^2}{n_j - 1}]}$$

where k = number of groups
n_j = size of the j^{th} group
S_j^2 = variance of the j^{th} group
$w_j = n_j / S_j^2$
$\bar{x}_j$ = mean of the jth group

$$\bar{x}_{adj.} = \Sigma \frac{w_j \cdot \bar{x}_j}{\Sigma w_j} \qquad \text{(adjusted overall mean)}$$

This F test has the regular *df* (*k*-1) for the numerator and an adjusted *df* for the denominator:

$$df = \{\frac{3}{(k^2-1)} \cdot [\Sigma \frac{(1-\frac{w_j}{\Sigma w_j})^2}{n_j - 1}]\}^{-1}$$

For testing HOV between two groups, O'Brien suggested that each contrast between two groups be conducted as a separate Welch ANOVA because the error term, MS_{error} "may be an inappropriate error term for specific contrasts…that do not involve all the cells of the design or have unequal absolute contrasting weights." (1981, p. 572) The researcher may control the Type I error rate by adjusting down the significance level through the Bonforroni method:

$$\alpha' = \alpha / K$$

where α = intended significance level for the study (usually 0.05)
α' = adjusted significance level for each contrast
K = number of contrasts

SAS statements to run Welch ANOVA are given below. Note the dependent variable SCORE refers to the transformed variable, not the original variable.

```
PROC GLM;
CLASS GROUP;
MODEL SCORE = GROUP;
MEANS GROUP / WELCH;
RUN;
```

Complex contrasts, e.g., Groups 1, 2 and 3 vs. Group 4 are possible using the same Welch ANOVA, according to O'Brien (1981). Therefore the O'Brien test can deliver more detailed analysis than any other methods, as the reader will soon see in IV. HOV Test for Factorial Designs.

As an example of the non-parametric alternatives, the **modified Sidney-Tukey test** (Conover, Johnson & Johnson, 1981) is explained here. It is not a widely adopted test, but it is interesting and practical enough to qualify for the list of selected HOV tests covered in this paper. The score of each person in the k groups is converted into an absolute deviation:

$$d = |x_{ij} - \bar{x}_j|$$

where x_{ij} = score of the i[th] person in the j[th] group
$\bar{x}_j$ = mean of the j[th] group

The absolute deviations are ordered from the smallest to the largest and assigned ranks in the following manner: the smallest d is ranked 1 and the largest d is ranked 2. In the remaining (N-2) ds, the largest gets 3 and the smallest, 4. In the next round of the remaining (N-4) ds, the smallest gets 5 and the largest, 6. A chi-square test is performed on the rank r.

$$\chi^2 = \Sigma n_j(\bar{r}_j - R)^2 / S_r^2$$

where n_j = size of the j[th] sample
$\bar{r}_j$ = mean rank of the j[th] group
R = mean of all the ranks
S_r^2 = unbiased variance of r

This modified Sidney-Tukey test has (k-1) degrees of freedom. Even though SAS does not list the option, the test can be run through other nonparametric methods under PROC NPAR1WAY, but the researcher need convert the scores first. In the SAS statements below, SCORE refers to the rank variable, r.

```
PROC NPAR1WAY;
CLASS GROUP;
VAR SCORE;
RUN;
```

SAS does not exactly perform the Sidney-Tukey test. When only two groups are involved, SAS runs the Wilcoxon rank sums test (Sidney & Castellan, 1988, pp. 128-137); with more than two groups, it switches to Friedman two-way analysis of variance by ranks (Sidney & Castellan, 1988, pp. 174-183). Their results are comparable to those of the modified Sidney-Tukey test. Because ranks, rather than absolute deviations, form the basis of the analysis, the test has less power. The reported χ^2 can be conveniently converted into F using the formula below:

$$F = [\, \chi^2 / (k-1) \,] \; / \; [\, (N - 1 - \chi^2) / (N - k) \,]$$

The F test has (k-1), (N-k) degrees of freedom. Type I error rate tends to be slightly higher when F approximation is adopted than when χ^2 is used. However, the difference is negligible (Conover, Johnson & Johnson, 1981, p. 360).

IV. HOV Test for Factorial Designs

For the two-way ANOVA fixed-effect factorial design, O'Brien proposed a robust procedure to test HOV (1979, 1981). The beauty of it is that it can attribute differences in variance to the main effects of independent variables A and B and the interaction effect AXB. It works with both balanced and unbalanced designs and allows subsequent multiple comparisons for more detailed analysis. It is the O'Brien test for one-way ANOVA generalized to the two-way situation. For the purpose of this paper, it is called the **generalized O'Brien test**, even though it is exactly the same test as the one explained above. The generalized O'Brien test is simply a two-way analysis of variance of the transformed variable, r, and Welch ANOVA can be conducted for pairwise comparisons with the significance level adjusted down through the Bonforroni method. The transformation to the spread variable r follows the formula:

$$r_{ijk} = \frac{(w + n_{jk} - 2)n_{jk}(x_{ijk} - \bar{x}_{jk})^2 - wS_{jk}^2(n_{jk} - 1)}{(n_{jk} - 1)(n_{jk} - 2)}$$

where w = weight (usually 0.5)
n_{jk} = size of the group at the j^{th} level of one independent variable and the k^{th} level of the other independent variable

x_{ijk} = score of the i^{th} person in the group at the j^{th} level of one independent variable and the k^{th} level of the other independent variable

$\bar{x}_{jk}$ = mean of the group at the j^{th} level of one independent variable and the k^{th} level of the other independent variable

S_{jk}^2 = variance of the group at the j^{th} level of the one independent variable and the k^{th} level of the other independent variable

SAS statements for two-way ANOVA with the transformed *r* variable SCORE as the dependent variable are listed below:

```
PROC GLM;
CLASS A B;
MODEL SCORE = A B A*B;
RUN;
```

O'Brien recommended Welch ANOVA for subsequent multiple comparisons (1981). The reader is referred to the discussion on the O'Brien test for the one-way ANOVA design for details.

V. HOV Tests for Two Correlated Samples

Correlated samples are typically involved in pre-post designs or studies that match the two subjects in each pair. HOV tests for such situations need to take into consideration the correlation between the two sets of scores. A positive correlation plus a statistically significant increase in variability indicates greater dispersion of prior differences. A positive correlation plus a statistically significant decrease in variability means reduction in prior differences. However, when a negative correlation occurs, the researcher may have to reconsider the research question and search for reasons other than the treatment to account for the reversal of the direction of individual differences. Should a zero correlation occur, the matching process has failed its purpose. The groups might as well be treated as independent samples. The discussion below proceeds on the assumption that the correlation is positive.

The **t-test** for the difference between the variance of two correlated samples is not available from SAS. Fortunately, it is simple enough for hand calculation. The t-test has (n-2) degrees of freedom.

$$t = \frac{S_1^2 - S_2^2}{[(1 - r^2)4S_1^2 S_2^2 /(n - 2)]^{0.5}}$$

where S_1^2 = variance under one condition
S_2^2 = variance under the other condition
r = correlation
n = sample size

A lesser known alternative to the above t-test is the **F_r test**, which follows the sampling distribution of the Pearson correlation r with df = (n-2) (Kanji, 1993, p. 38). First the ratio of the larger variance to the smaller variance is calculated (F′). Then F_r is computed using the following formula:

$$F_r = (F' - 1) \ / \ [(F' + 1) - 4r^2F']^{0.5}$$

where F′ = variance ratio
r = correlation

Critical values for various degrees of freedom at the 0.05 or 0.01 level of significance are available from the Pearson correlation table in most statistics textbooks.

It may be in order here to call the reader's attention to the possibility of extending Levene's test to the pre-post design of testing HOV (Rosenthal & Rosnow, 1991, p. 340), that is, conducting the regular repeated measures ANOVA on the absolute deviations. But this author is not aware that the procedure has been validated through mathematical proofs or *Monte Carlo* studies. Should such an approach prove to be feasible, it might have very interesting implications for the largely unknown territory of HOV testing involving more than two repeated measures.

VI. Summary

The 14 tests discussed in the paper are representative of four major approaches to HOV testing. The major approaches outlined below may serve as an efficient mental organization for nearly all the HOV tests. Most of them have not been included in this paper because they are judged, in comparison to the selected 14 methods, to be redundant, inaccurate or too elaborate to be practical for applied research.

The conceptually most straightforward major approach deals directly with the variance, or more frequently, the variance ratio, e.g., the one-sample χ^2 test, two-sample folded form F test, Hartley's F_{max} test with David's multiple comparison procedure, Cochran's G test, t-test for two correlated samples and F_r test for two correlated samples. Unfortunately, this approach is also most sensitive to symmetry and kurtosis. Those tests are easy but not robust. Many of those tests cannot deal with unbalanced designs. These tests are most likely to be mentioned in introductory level statistics or research design textbooks often without the caveat that they represent the least robust approach to HOV testing.

The second major approach relies on the natural log transformation of the variance because the log variance approximates the normal distribution quite well. The Bartlett-Kendall test and Bartlett χ^2 test in this paper demonstrate the approach. Likelihood ratio tests based on log variance are more robust than variance ratio tests. However, many statisticians still feel that they are quite vulnerable to deviations from normality.

The third approach applies the logic of ANOVA to transformed variables. Tests, such as Levene's test, Brown-Forsythe test, and O'Brien test with Welch ANOVA serving as a prudent procedure for multiple comparisons, have a strong appeal to non-statistician researchers and compare favorably to all the other approaches in terms of power and robustness. Among the three tests discussed in this paper, the Brown-Forsythe test and the O'Brien test may have overall advantage over Levene's test. The O'Brien test is particularly appealing because it applies to both one-way and two-way ANOVA and comes with a handy Welch-type procedure for multiple comparisons, all of which can be accomplished within SAS/STAT. Methodologically, it is also more sophisticated because it allows kurtosis to come into play through the weight w. This author recommends the ANOVA approach for a pedagogical reason as well. Since HOV is typically discussed in conjunction with ANOVA, ANOVA on differences among means and ANOVA on differences among variances share the same logic. Directing student's attention to HOV ANOVA serves to reinforce the conceptual understanding of ANOVA and at the same time addresses the issue of heterogeneity of variance, a topic largely ignored in most of the textbooks.

The last major approach to HOV testing is the nonparametric alternative represented by the modified Sidney-Tukey test. In the past, many attempts were made to conduct HOV testing by way of ranks to simplify computation. All of them use the chi-square approximation. With the easy access to computers today, those methods do not seem to have much to recommend themselves for, and they are not available from most of the software packages. Attention to the nonparametric alternatives has been declining. It is quite possible that those methods will eventually be replaced by the ANOVA approach.

References

Bartlett, M.S., & Kendall, D.G. (1946). The statistical analysis of variances - heterogeneity and the logarithmic transformation. *Journal of the Royal Statistical Society, Supplement 8*, 128-138.

Canover, W.J., Johnson, M.E., & Johnson, M.M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, *23*, 351-361.

David, H.A. (1954). The ranking of variances in normal populations. *Journal of the American Statistical Association, 51*, 621-626.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62*, 61-84.

Ferguson, G.A. (1981). *Statistical analysis in psychology and education* (5th ed.). New York: McGraw-Hill.

Gould, S.J. (1996). *Full house: The spread of excellence from Plato to Darwin.* New York: Three Rivers Press.

Hedges, L.V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research, 63*, 94-105.

Kanji, G.K. (1993). *100 statistical tests*. Newbury Park, CA: Sage Publications.

Noddings, N. (1992). Variability: A pernicious hypothesis. *Review of Educational Research, 62*, 85-88.

O'Brien, R.G. (1979). A general ANOVA method for robust tests of additive models for variances. *Journal of the American Statistical Association, 74*, 877-880.

O'Brien, R.G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin, 89*, 570-574.

Pearson, E.S., & Hartley, H.O. (1970). *Biometrika tables for statisticians*. Cambridge, UK: Cambridge University Press.

Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

SAS Institute (1997). *SAS/STAT software: Changes and enhancements through Release 6.12*. Cary, NC: SAS Institute.

Shaffer, J.P. (1992). Caution on the use of variance ratios: A comment. *Review of Educational Research, 62*, 429-432.

Sidney, S. & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Tietjen, G.L. & Beckman, R.J. (1972). Tables for use of the maximum F-ratio in multiple comparison procedures. *Journal of the American Statistical Association, 67*, 581-583.

Welch, B.L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika, 38*, 330-336.

Winer, B.J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Fourteen Homogeneity of Variance Tests: When and How to Use them

Author(s): Shuqiang Zhang

Corporate Source: Department of Educational Psychology
University of Hawaii at Manoa
Honolulu HI 96822

Publication Date: Presentation at '98 AERA
4/16/1998

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Shuqiang Zhang

Printed Name/Position/Title: SHUQIANG ZHANG / Associate Professor

Organization/Address: Department of Educational Psychology
University of Hawaii at Manoa
1776 University Avenue, Wist 208
Honolulu HI 96822

Telephone: (808) 956-4295

FAX: (808) 956-6615

E-Mail Address: szhang@hawaii.edu

Date: 4/24/1998

(over)